Data Mining Based Study to Analyze Passenger Transportation in Uva Province

W.B.M.S.C. Wijayakoon
Department of Information and Communication Technology
Uva Wellassa University of Sri Lanka
Badulla, Sri Lanka
supunw@uwu.ac.lk

W.M.C.J.T. Kuthulwatta
Department of Information and Communication Technology
Uva Wellassa University of Sri Lanka
Badulla, Sri Lanka
chirantha@uwu.ac.lk

J.K.H. Samapth
Department of Biosystems Technology
Uva Wellassa University of Sri Lanka
Badulla, Sri Lanka
hasitha@uwu.ac.lk

Kolitha B. Wijesekara
Department of Biosystems Technology
Uva Wellassa University of Sri Lanka
Badulla, Sri Lanka
kolitha@uwu.ac.lk

***Abstract*— Passenger transportation methods are a very important factor for sustainable development and sustainable supply chain management. According to the pandemic situation and to employ well established economic and social based development within the Uva Province, there is a need of identifying any data patterns, uncovered information, and obtaining other valuable information. The primary objective of this research is to uncover patterns, correlations, and trends that influence transportation operations. According to the several studies, data mining techniques were used for data analysis, pattern identification and decision making. Hence, this research analyses road passenger transportation within the Uva Province in Sri Lanka using data mining techniques. The dataset of 26,100 entries, managed by the Passenger Transport Service Authority of Uva Province from 1976 - 2022, was collected and explored to uncover patterns, correlations, and trends that influence transportation operations. This analysis reveals several significant findings. The findings contribute to evidence-based recommendations aligned with current transportation trends and market conditions. This paper underscores the significance of data mining in transforming raw data into actionable insights that facilitate informed decision-making in the transportation sector.**

***Keywords*—** *Data mining, Passenger Transportation, Analytics, Transportation patters*

I. INTRODUCTION

In light of the pandemic and the need for effective economic and social development in Uva Province, it's essential to identify data patterns and uncover valuable insights. The main goal of this research is to reveal the patterns, correlations, and trends that impact transportation operations.

Data mining has been called exploratory data analysis, among other things. Masses of data generated from cash registers, from scanning, from topic-specific databases throughout the company, are explored, analysed, reduced, and reused. Searches are performed across different models proposed for predicting sales, marketing response, and profit. Classical statistical approaches are fundamental to data mining.

Risk management, data security planning, and many other crucial business use cases all greatly benefit from data mining. It is crucial in a variety of other professions as well, including governance, sports, mathematics, and scientific research. Patterns, correlations, and trends will be found in this study's data collection on the transportation services under the control of the Passenger Transport Service Authority of Uva Province (PTSAUP), and recommendations will be made in accordance with forecasts based on transportation trends, behaviours, and market conditions discovered using data mining techniques.

All private bus transit is governed and observed by the Road PTSAUP (RPTAUP). The Ministry of Sports, Youth Affairs, Tourism, Transport, Cultural Affairs, Textiles, and Small Industries is in charge of the RPTSAUP. According to Sections No. 01 of 2001 and No. 02 of 2019 (revised) of the Statute of the Passenger Transport Service Authority, RPTSAUP was created in the year 2000. The goal was to efficiently provide passenger transport services. The RPTSAUP continues to provide services for the administrative districts of Badulla and Monaragala and maintains positive working relationships with the national Ministry of Transport and all other relevant public and private institutions in order to carry out the government's policies for the development of the transportation sector. To meet the public transportation needs of the Uva Province, 12 and 14 bus stations available in Monaragala and Badulla district respectively where 26 in total. Main research objectives of this research are as follows:

- RO1 - To discover any structure in the unstructured data set based on RPTSAUP.
- RO2 - To extract any meaning from noisy data based on RPTSAUP.
- RO3 - To discover any patterns in apparently random data on the data set.
- RO4 - To understand any patterns, correlations, and trends on the selected data set.
- RO5 - To predict any transportation trends, behaviours, and market states.

II. Literature Review

The practice of drawing insightful conclusions from huge databases is known as data mining. It has been widely used in a variety of industries, including transportation, where it is used to examine trends and behaviours in the industry.

To find a solution in the real world of transportation, numerous pieces of information must be combined. Decision-makers for transportation engineers now have additional options thanks to recent studies in transportation data mining. K-means clustering [1], decision trees [2], and neural networks [3] were used to forecast freeway trip time with non-recurrent congestion [4].

Data mining techniques such as clustering and classification were utilized in public transportation data to analyse passenger travel behaviour [5]. Similar study used clustering analysis to group passengers based on travel patterns and identified key factors affecting passenger travel behaviour. The study found that data mining could be an effective tool for bettering public transportation services and comprehending passenger behaviour [6].

In a different study, researchers employed clustering techniques to discover a relationship between the route taken by travellers and the flow of passengers at the station [7]. Furthermore, in [8] authors collected data in a metro system to estimate passenger route choice patters where probabilistic model was developed to estimate from empirical analysis. Six phases proposed in [9,10] that can be taken into account for data mining initiatives, according to the Cross-Industry Standard Procedure for data mining is categorized as, Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment as shown in Fig. 1.

In the field of transportation engineering, a lot of information has been gathered about traffic, accident analysis, asphalt conditions, road durability and stair management. It is identified that bus optimization and predict number of passengers in [11] used MARS algorithm to predict optimal demand. Random Forest model used in [12] to better predict regression. Furthermore, Naïve bayes model to forecast bus demand [13] and LSTM forecasting model [14] to predict short-term public transportation demand.

In conclusion, it has been demonstrated that data mining is an effective technique for examining trends and behaviours in transportation data. This study highlights the need for research in this area and offers additional proof of the utility of data mining in the transportation sector.

III. Methodology

Our stated research methodology is to find

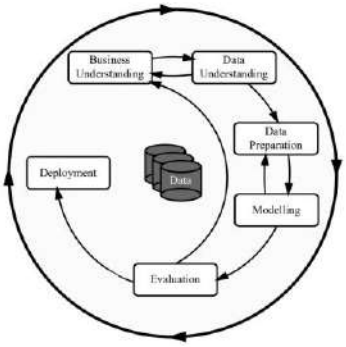


Fig. 1. Data mining model [10]

structure in chaotic data, from noisy data to find patterns and connections, and predict trends and behaviors. Mainly focused on the most recent data mining techniques put forth by researchers and consult

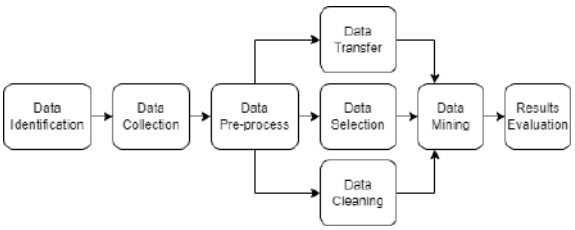


Fig. 2. Methodology

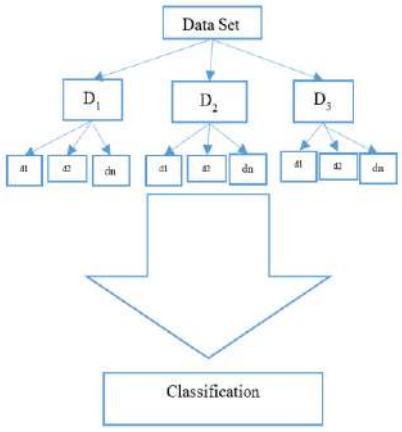


Fig. 3. Data classification steps

recently published research articles to gain the necessary knowledge to create well-organized research. As the main research steps in the methodology as shown in Fig. 2, data from the

RPTSAUP was identified, collected and preprocessed. Then applied data classification and clustering techniques. Furthermore, ZeroR classifier was used to visualize patterns and trends available in data.

A. *Data Collection, Pre-Processing and Data Mining*

Initially, a dataset of 26,100 entries was collected from the RPTSAUP at Uva Provincial Council and translated from Sinhala language to English language. Next, null values were cleaned, followed by the classification of the data. Data classification steps were illustrated in Fig. 3. These data are mainly classified into three parts as, bus details (8100 data), bus owner details (9000), and bus driver and conductor details (9000). Using the number of bus seats in the Bus Details data set, buses were divided into three types. These types are one-door buses, two-door short buses, and two-door long buses. They were divided into Badulla and Monaragala districts according to the locations of bus stands and divisional secretariats. The bus driver's and conductor's ID numbers were used to determine their age limits.

IV. RESULTS AND DISCUSSION

A. *Bus Details Based on Year of Manufacture*

The bus fleet of PTSAUP bus operations consists of over 1016 buses from different manufacturing years which range between 1976 to 2022.

According to Fig. 4, highest number of the buses in Uva Province were manufactured in 2011, which is 132. Majority of the buses were manufactured between 2000 and 2018. There are very few buses used which manufacture year is 2019 onwards.

B. *Analysis of Bus Details based of the Number of Seats and Bus type:*

PTSAUP Bus Operations bus fleet comprises over 1016 buses, from different manufacturers. All buses are 102 in. (2.59 m) wide. Most of these bus models consist of double seats on each side. The driver's seat is isolated and the back row usually has more seats than others. The typical bus seat is 39 inches wide and generally considered to have a maximum seating capacity of three. An average bus carries between 30 and 100 passengers.

Fig. 5 shows the number of seats in buses in Uva province. Most of the buses in Uva Province have 42 seats. It is two doors (short) busses. 303 buses belong to it. There are also 248 buses with 54 seats. It is two doors (long) busses. There are fewer single-door buses as compared to other buses.

There are three types of buses in Uva Province as one-door buses, two-door short buses, and two-door long buses. The most common type is the two doors (short) bus. A conventional door, also known as a regular door is a type of door that is hinged at the front-facing edge of the door, and so

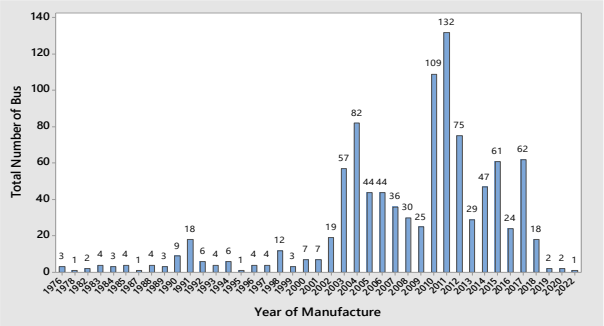


Fig. 4. Representation of buses by year of manufacture

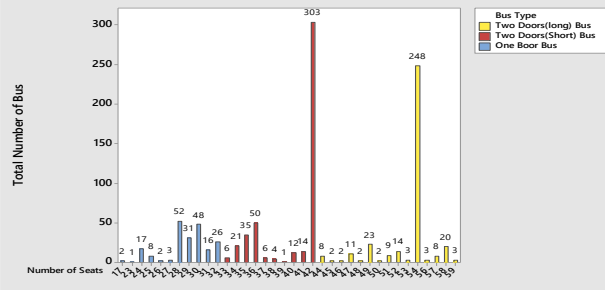


Fig. 5. Number of bus count based on bus type and number of seats available

allows the door to swing outward from the hull of the bus. These doors are relatively safe, in that if they are opened during forward motion of the vehicle, the wind resistance will work against the opening door, and will effectively force its closure. The purpose of having two doors on a bus is that it saves time at stops if boarding passengers don't have to wait for those alighting to get off.

Fig. 6 shows the type of buses. This graph clearly shows the greatest number of buses is two doors (short) buses.452 buses belong to it. There are also 358 buses with two doors (long buses). There are also 206 buses with one door in Uva province. In general, bus types are distributed as 20% are one-door buses, 45% are short buses, and 35% are long buses.

C. *Analysis of Bus Details based on the Provided Service:*

Buses were clustered in to 6 based on the service that they provide. They are Air-Conditioned Buses (AC), Normal Service (N), Normal and Sisuseriya (N/S), Waiting Buses (P), Nisiseriya Service (Nisi S), Sisuseriya (S).

Fig. 7 shows the clustering of bus data based on the service provided by the bus. This clearly shows that the majority of them were normal service providers which is 88%. There were 889 normal service providers, 61 Sisuseriya buses, 30 AC buses and 29 Normal /Sisuseriya buses in Uva Province. However, there were 6 Waiting buses in uva province and one bus to Nisiseriya service.

D. *Analysis of Data Based on Bus Stands*

A bus stand is typically used to allow a bus to stop at a bus stop. Additionally, bus stops permit

short-term parking for driver changes or breaks. A bus stand is the point where a bus route starts or ends, where they stop, turn or reverse, and wait before departing on their return journeys. It's also a place for passengers to board and alight from vehicles. It also often provides a convenient point where services can be controlled from.

Fig. 8 shows the bus stops of existing buses in Uva Province. At present, transport service facilities are being provided

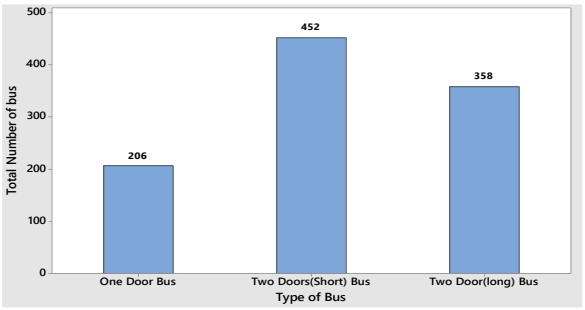


Fig. 6. Total number of bus count based on bus type

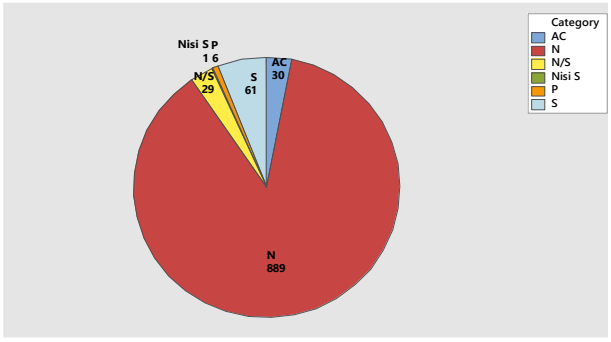


Fig. 7. Number of bus count based on the providing service

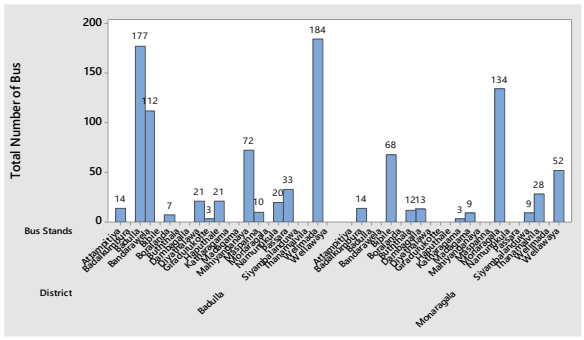
through 26 bus stations in both Badulla and Monaragala districts to fulfil the passenger transport public service needs of Uva Province. Welimada bus station has 184 buses, making it the bus station with the most buses. Next to that, Badulla, Monaragala, Bandarawela, Mahiyanganaya, and Bibile bus stations contains the most buses. The number of buses belonging to them are 177, 134, 112, 72, and 68, respectively. Out of 26 bus stations in Uva province, buses were registered for only 22 bus stations.

E. Divisional Secretariat Division (DSD)

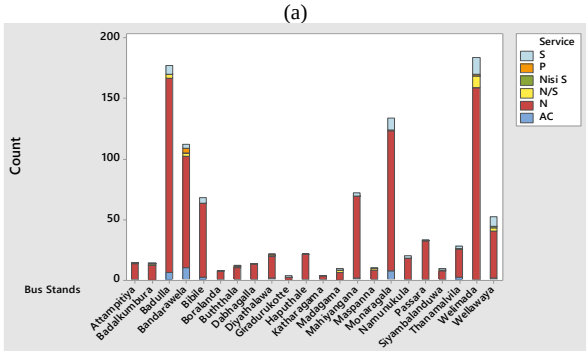
Divisional secretariats are the administrative subdivisions that make up Sri Lanka's districts. These were initially based on the *korales* and *ratas*, two types of feudal counties. Previously, they were referred to as "D.R.O. Divisions" in honour of the "Divisional Revenue Officer". Later, the D.R.O.s changed their title to "Assistant Government Agents", and the Divisions adopted the name "A.G.A. Divisions". The divisions are currently managed by a "Divisional Secretary" and are referred to as "D.S. Divisions". In Uva Province, there were 27 DSDs identified by district. Among them, 11 in the Monaragala district and 16 in the Badulla district.

Based on Fig. 9, DSD of Welimada has the highest number of buses which is 103 and Lunugala DSD has the least number of buses which is 3.

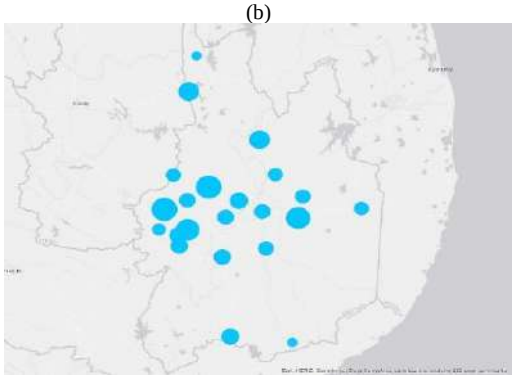
According to the Fig. 10, highest number of bus count registered under Badulla district holding 66% from total number of buses. Out of 1016 buses registered under the PTSAUP, 674 buses belong to the Badulla district, and 342 buses belong to the Monaragala district.



(a)

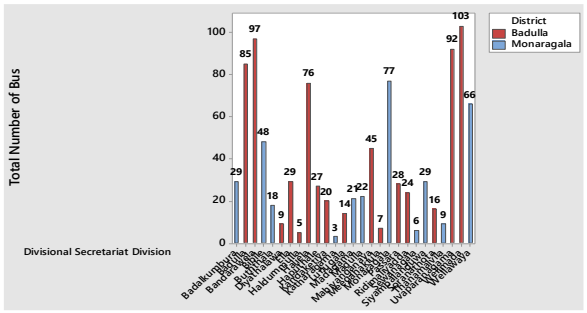


(b)

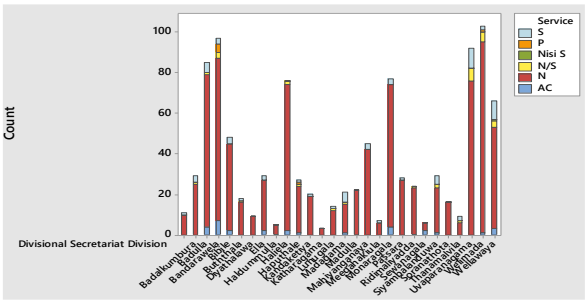


(c)

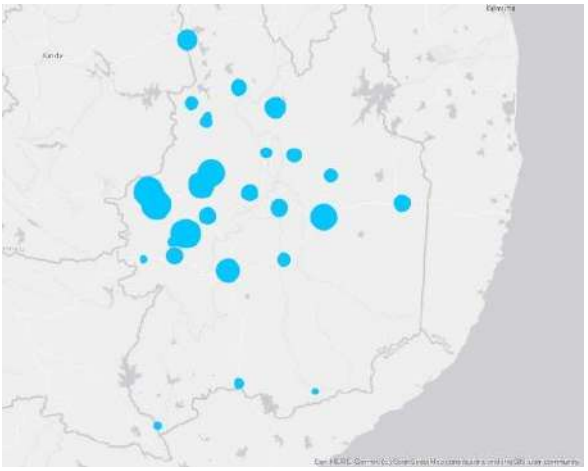
Fig. 8. (a) total number of buses registered for bus stands, (b) total number of buses registered for bus stands with their service. (c) mapping of bus stands with the number of buses



(a)



(b)



(c)

Fig. 9. (a) total number of buses registered under DSDs and relevant district, (b) total number of buses registered under DSDs and service, (c) mapping of bus DSDs with the bus count

F. *Bus Routes and Distance Traveled Per Day*

Fig. 12 (a) shows the routes travelled by buses belonging to PTSAUP. Thickness of the route is generated based on the number of buses travelling along the route. Fig. 12 (b) shows the number of times (trips) buses travel per day. Maximum number of buses travel two trips per day. Furthermore, there is a bus that travels 12 trips per day which is the highest. Details of that bus is shown in Fig. 11. It is a normal service bus, and its route is Halieala to Badulla and this bus belongs to the Halilela DSD.

G. *Analysis of Bus Owner Details*

Among all bus owners majority possess one bus. Very few owners possess more than one bus. However, maximum number of buses that owns by one person is 5. Out of those five, four were served as Sisuseriya and one served as Normal Service. Fig. 13 represents the information of those 5 buses.

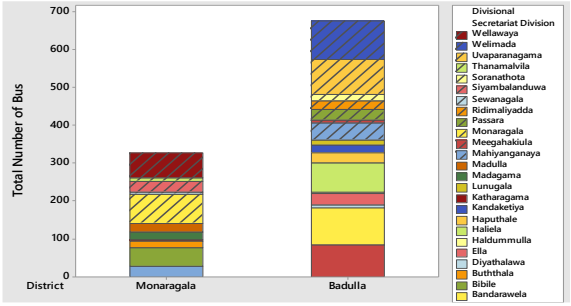


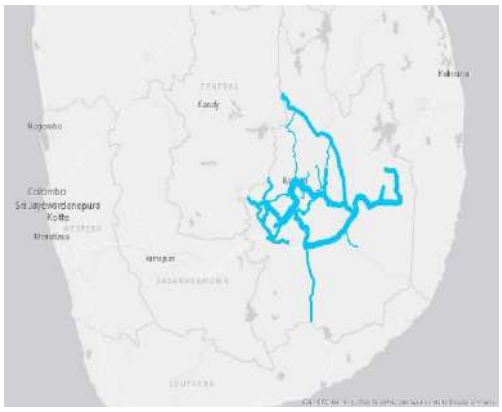
Fig. 10. Representation of the relationship between number of buses and bus stands registered under two districts

H. *Place of departure and destination*

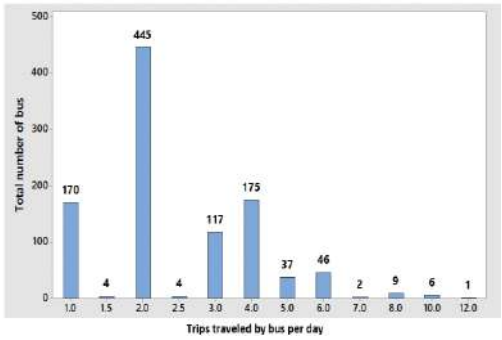
Place of departure is the point where a bus route starts and where wait before departing on their return journeys. It's also where passengers board buses. A bus place of destination is the point where a bus route ends, where buses stop or turn. It's also where passengers alight from buses. Badulla and Bandarawela are the departure and destination points for most buses in the PTSAUP. The majority of buses departure from Bandarawela and end the trip at Badulla as shown in Fig. 14 and Fig. 15.

I. Bus Drivers and Conductors Details

1) Bus Drivers: Bus drivers transport people between diverse locations, such as work, school, and towns.



(a)



(b)

Fig. 11. (a) Route mapping based on the number of buses travelled, (b) Distance travelled by bus per day

Maximum trips traveled by bus per day = 12

Divisional Secretariat Division	Haliela
Road No	296
Distance (Km)	6.40
Trips traveled by bus per day	12
Route approved to run	Haliela - Badulla
Service	N
Bus Stands	Badulla

Fig. 12. Details of the bus travelled maximum trips per day

They follow a daily schedule while transporting people on regular routes along town or suburban streets. They stop frequently, often every few blocks and when a passenger requests a stop.

Fig. 16 shows the number of drivers registered in PTSAUP based on age ranges. There were 337 drivers in the age range between 34 and 43 years where it holds the highest group of drivers while 13 drivers fall in the age range from 64–73 years where it holds the lowest group of drivers.

2) Bus Conductors: Bus conductors have a range of responsibilities, which include collecting fares, suing tickets, helping passengers with baggage, lvising on destinations and ensuring the safety of assengers.

Fig. 17 shows the number of bus conductors gistered in PTSAUP based on age ranges. There ere 265 conductors in the age range of 24–33 and it the largest group of conductors among all age nges. There were 8 conductors in the age range of 4–73 and that is the smallest group of conductors.

V. CONCLUSION

Based on the results from the analysis of passenger ansportation data in Uva Province, offers valuable sights and suggestions for improving the efficiency d effectiveness of bus operations. The majority of ses in service were found to be relatively modern hich were manufactured between 2000 and 2018. It better to continue bus upgrade efforts to enhance the rvice quality and passenger comfort. eographically, Badulla district becomes the hub for s operations while having the highest ncentrations of buses, especially in the Welimada

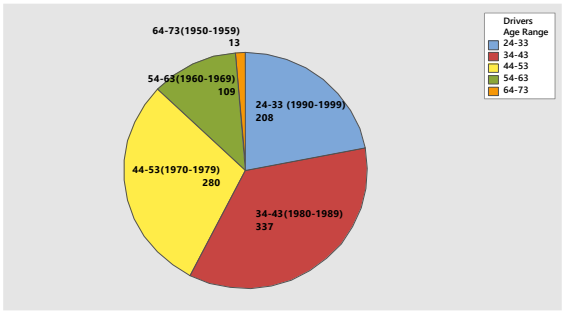


Fig. 16. The number of drivers based on age range

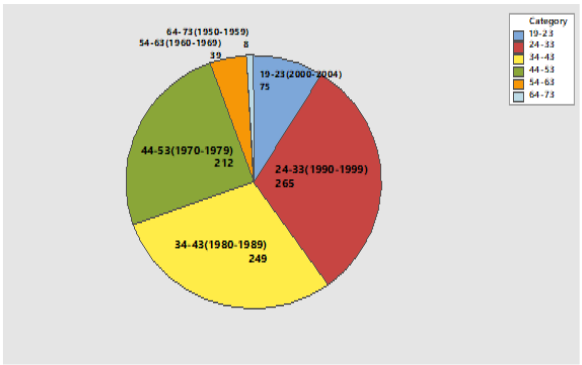


Fig. 17. The number of conductors based on age range

sector in Uva Province to improve public transportation services and enhance the overall transportation system.

The application of data mining techniques was facilitated by Minitab and Weka software to enable robust analysis. Furthermore, ArcGIS software was used for bus route mapping from patterns obtained by data clustering and operational patterns by demonstrating the potential of advanced analytics in enhancing public transportation management and service delivery.

ACKNOWLEDGMENT

We would like to express our gratitude to the Uva Wellassa University of Sri Lanka for the financial support of this research (UWU/RG/2022/19) and the Road Passenger Transportation Authority of Uva Province for providing required data and information.

REFERENCES

[1] R. A. Kadir, Y. Shima, R. Sulaiman, and F. Ali, "Clustering of public transport operation using K-means," in 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), 2018, pp. 427-432.

[2] J. de Oña, R. de Oña, and G. López, "Transit service quality analysis using cluster analysis and decision trees: a step forward to personalized marketing in public transportation," Transportation, vol. 43, pp. 725-747, 2016/09/01 2016.

[3] D. Shmueli, "Applications of neural networks in transportation planning," Progress in Planning, vol. 50, pp. 141-204, 1998/10/01/ 1998.

[4] C.-S. Li and M.-C. Chen, "A data mining based approach for travel time prediction in freeway with non-recurrent congestion," Neurocomputing, vol. 133, pp. 74-83, 2014/06/10/ 2014.

[5] Z. Hu, L. Jingen, and H. Bing, "Analysis of passenger travel behavior based on public transportation OD data," presented at the Proceedings of the 6th International Conference on Information and Education Technology, Osaka, Japan, 2018.

[6] X. Cao, H. Chen, F. Liang, and W. Wang. (2018, Measurement and Spatial Differentiation Characteristics of Transit Equity: A Case Study of Guangzhou, China. Sustainability 10(4).

[7] N. Lathia, J. Froehlich, and L. Capra, "Mining Public Transport Usage for Personalised Intelligent Transport Systems," in 2010 IEEE International Conference on Data Mining, 2010, pp. 887-892.

[8] J. Zhao, F. Zhang, L. Tu, C. Xu, D. Shen, C. Tian, et al., "Estimation of Passenger Route Choice Pattern Using Smart Card Data for Complex Metro Systems," IEEE Transactions on Intelligent Transportation Systems, vol. 18, pp. 790-801, 2017.

[9] G. S. Linoff and M. J. A. Berry, Data Mining Techniques. For Marketing, Sales and Customer Support: Wiley Computer Publishing, 1997.

[10] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," 2000.

[11] M. F. Khan, S. Asghar, M. I. Tamimi, and M. A. Noor, "Multi-Objective Transport System Based on Regression Analysis and Genetic Algorithm Using Transport Data," IEEE Access, vol. 7, pp. 81121-81131, 2019.

[12] C. Xue, Y. Ju, S. Li, and Q. Zhou. (2020, Research on the Sustainable Development of Urban Housing Price Based on Transport Accessibility: A Case Study of Xi'an, China. Sustainability 12(4).

[13] T. Stadler, A. Sarkar, and J. Dünnweber, "Bus Demand Forecasting for Rural Areas Using XGBoost and Random Forest Algorithm," in Computer Information Systems and Industrial Management, Cham, 2021, pp. 442-453.

[14] J. D. Caicedo, M. C. González, and J. L. Walker, "Public Transit Demand Prediction During Highly Dynamic Conditions: A Meta-Analysis of State-of-the-Art Models and Open-Source Benchmarking Infrastructure," arXiv preprint arXiv:2306.06194, 2023.